Visualizing Industrial Dynamics via Stock Market

Visualizing Industrial Dynamics Through Represented Stock Capital Changes Over about the Past Decade.

Chun Yuan
u6052162@utah.edu, u6052162
Zixuan Lu
birdpeople1984@gmail.com, u6053162

Content

20 | Data Processing

22 | Exploratory Data Analysis and Design Evolution

27 | Implementation

32 | Final results and evaluations

This page is intentionally left blank.

Project Proposal

1.1 Basic info

1. **Project title:** Visualize Dynamics of Stock Market
2. **Group members:**
   Chun Yuan, u6052162@utah.edu, u6052162;
   Zixuan Lu, birdpeople1984@gmail.com, u6053162.
3. **Repo:** https://github.com/dataviscourse2024/group-project-market-dynamic-visualization.

1.2 Background & Motivation

Analyzing the industrial structure offers an effective perspective for investment analysis. Over time spans ranging from several decades to a century, countless emerging industries have been born and have flourished, while older, less competitive industries have been phased out or downgraded. Observing and analyzing these trends provides valuable insights, allowing for predictions of economic cycles and offering macroeconomic investment guidance.

Industries are always evolving, driven by technological advances, consumer behavior, and regulations. New industries, like digital technology, renewable energy, and biotechnology, have created significant growth opportunities by challenging older business models and reshaping markets on the most recent decades. These sectors can offer high returns to investors who recognize and invest in them early. On the other hand, older industries that fail to adapt may decline, as seen with traditional manufacturing in many developed countries. Understanding which industries are likely to struggle can help investors adjust their portfolios to avoid potential losses.

By systematically monitoring these industrial shifts, investors can develop a more nuanced understanding of economic cycles. They can identify emerging opportunities and potential risks, and allocate their capital more efficiently. Such analysis is essential for formulating long-term investment strategies that align with evolving economic realities.

1.3 Project Objectives

In this project, we aim to provide an effective method for visualizing the dynamics of industrial structures. The major questions we wants to answer is:

1. What are the general patterns of the market (economic cycle patterns)?
2. Do industries developments reflect economic cycle patterns (cyclical/counter-cyclical industry)?
3. Are industreis growth balanced over a limited period?

4. Are there significant events that have had observable impacts on certain industries?

We believe these four questions are crucial for understanding how industries perform relative to the overall economic cycle at a limit time-span and for making informed investment decisions.

Considerations for Project Execution and Progress:

Note. Rather than creating our own classification standards, we will use the **Global Industry Classification Standard (GICS)**, which is widely adopted in the industry[1, 2]. There are numerous economic indicators that can reflect changes in industries, but we want to keep our approach straightforward and focused. Given that **stock market capitalization** is a direct indicator of industrial trends, preferences, and the development of companies and industries, we will use the **market capitalizations of a few representative stocks within each sector** as our key metric. This will allow us to visualize how their market values have changed over a certain period.

Since this is a short-term project designed for a single semester, we will limit the time frame of the data we visualize to **the most recent 5-10 years**, essentially covering up to a decade. This time frame will provide us with a manageable yet meaningful period for observing significant trends and shifts within various industries.

By focusing on a concise time span and using a clear, widely recognized standard like GICS, we believe our visualization will offer a practical tool for understanding industry dynamics without overwhelming complexity.

1.4 Data

1.4.1 GICS Industrial Sector and Industry

The Global Industry Classification Standard (GICS) is a widely recognized system for categorizing companies into specific sectors and industries, developed by MSCI and S&P Global in 1999, providing a consistent framework for analyzing companies worldwide, allowing investors and analysts to compare performance and trends across various sectors and industries. It divides the global market into 11 sectors, 24 industry groups, 69 industries, and 158 sub-industries, covering the full spectrum of economic activities.

For this project, we will use GICS to classify companies and industries with sector and industry tag. To be specific, we will select the stock based on this classification. The stock will be clustered respected to such tag. A group of stock within the same industrial tag will be grouped up to reflect the development of such industry. We will use 2-level of the classification information, which means specify to GICS industry group (omitting industies-level and sub-industries-level). Such classification information can be found at GICS white paper book.

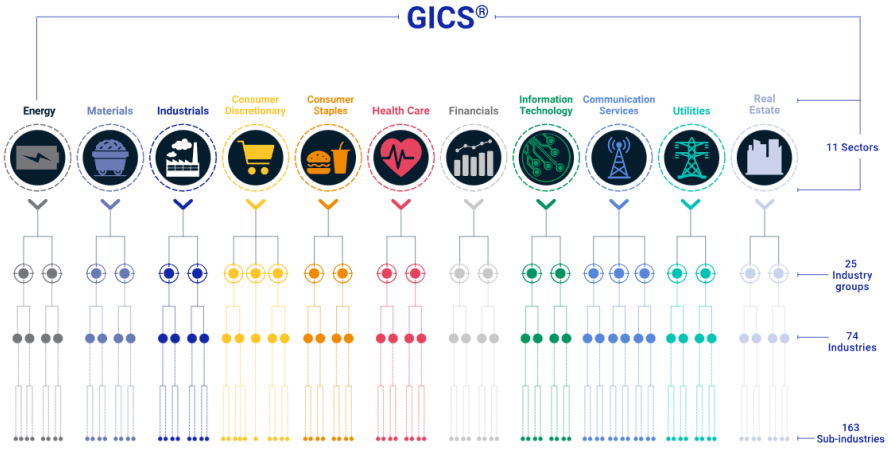


Figure 1.1: GICS Overview.

1.4.2 Stock Code with Tag

On the next stage, we will obtain the stock code list based on GICS Industrial sector and industry, which means we will select a group of represented stock with stock code based on GICS classification. Fortunately, we found a public data based on S&P 500 (2022), which collects totally 512 stock with {Ticker Exchange, GICS Sector, GICS Industry Group, GICS Industry, GICS Sub-Industry} as meta-data. This data can be found at Here.

Ticker	Exchange	GICS Sector	GICS Industry Group	GICS Industry	GICS Sub-Industry

Figure 1.2: Attribute: stock code based on GICS.

1.4.3 Stock meta-data

Finally, with the tagged stock code list, we will query a public database to obtain the meta-data of each single stock. The meta-data we queried for each stock will contain closing pric, market capitalization, trading volume, etc, combined with its stock code and GICS tag.

The public open source stock meta-data database we are going to query is: AKShare. We will use 5-10 years (end for 2024.8) with week-level interval data. This meta-data has to be query within Python. On the next section we will show the data processing strategy to format our data.

1.5 Data Processing

We will collect and manipulate our data to form a final format. Each entry of the data will have these attributes:

1. Time (quantity): date of the stock
2. Stock code (category): the code of the represented stock
3. Closing price (quantity): the closing price of the stock
4. Market capitalization (quantity): the market capitalization of the stock
5. Trading volume (quantity): the trade volume of the stock

6. GICS sector tag (category): one of the 11 GICS sector tag

7. GICS industry tag (category): one of the 25 GICS industry tag

1.6 Visualization Design

1.6.1 Overview

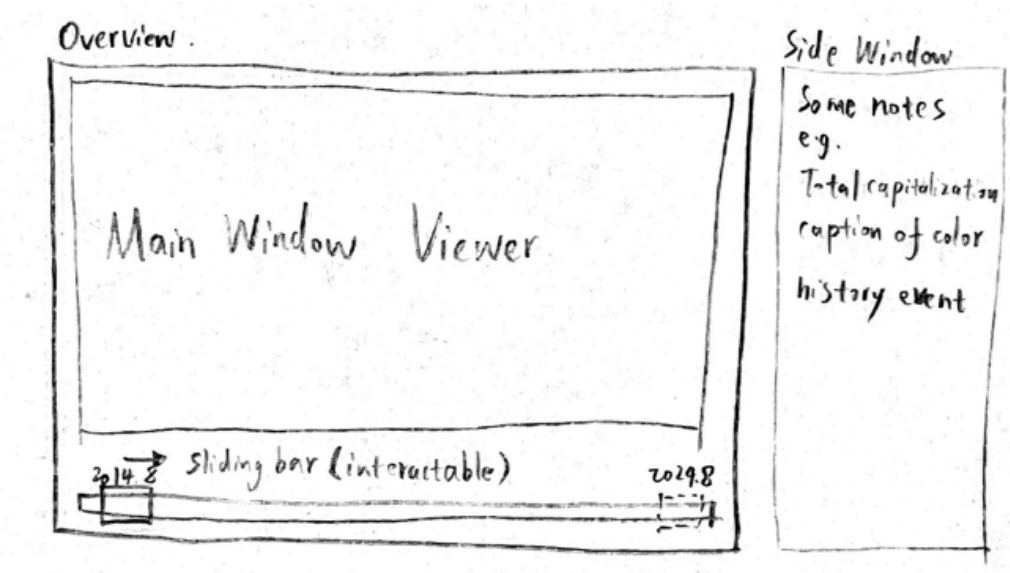


Figure 1.3: Overview of the visualization.

Figure 8.2 provides an overview of our visualization interface. The main interface consists of a side panel for displaying notes, an interactive sliding bar (timeline), and a central viewer for visualization. As the user adjusts the timeline slider, the content in the main viewer dynamically updates to reflect the selected time frame. The side panel presents relevant text, such as important historical events and other contextual information.

1.6.2 Specific design on main windows

The main window will display stock prices corresponding to each GICS sector or industry to illustrate the evolution of a given industry. It will present both the performance of individual stocks and the total market capitalization of the entire industry. To achieve this, we have developed three potential solutions:

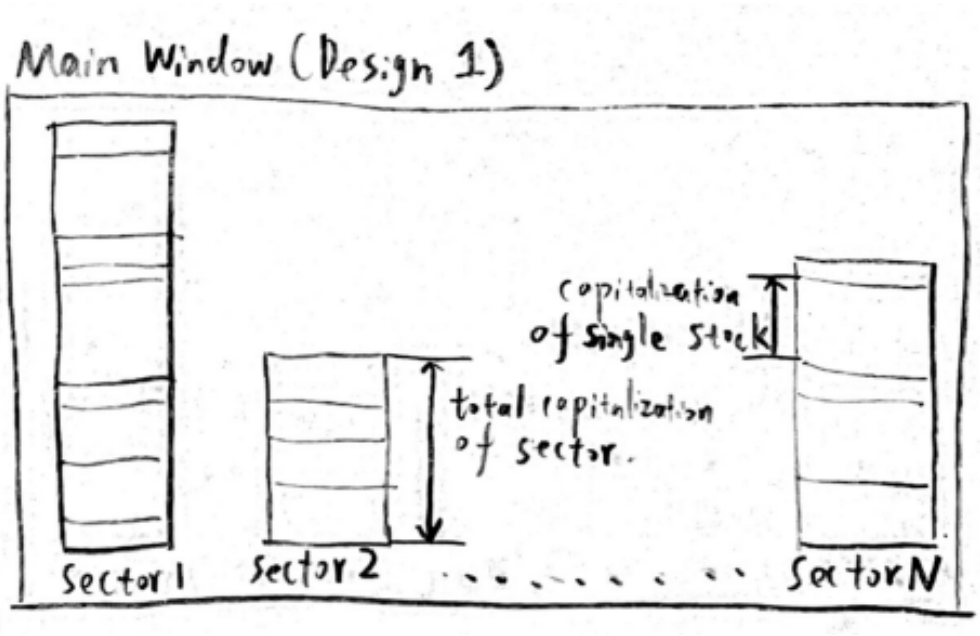


Figure 1.4: Design 1: Using stacked bar diagram.

The first idea we considered was using a traditional stacked bar chart(Fig. 1.4) to visualize the data. However, this approach was discarded for two main reasons. First, if a single industry grows rapidly, the axis may change significantly, leading to an axis-compression problem, where a large maximum scale can obscure smaller values. Second, each industry or sector may include stocks of varying sizes, which can create issues when stacking them on a unified scale. Despite these drawbacks, this design does effectively display both the total market capitalization and the values of individual stocks.

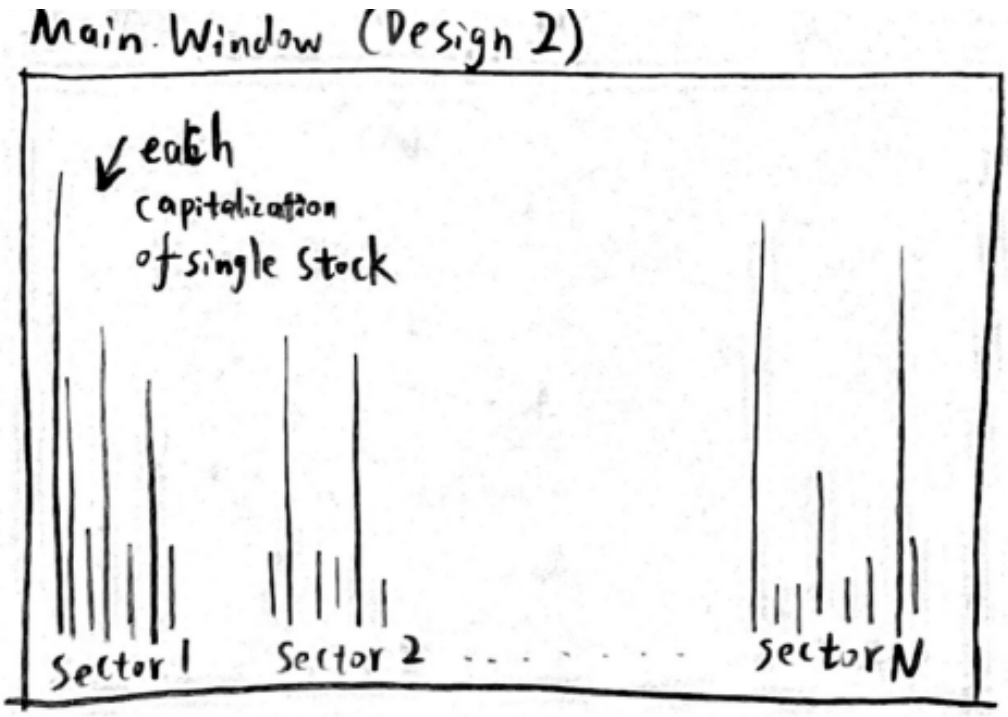


Figure 1.5: Design 2: Using separated bar diagram.

Another idea is to visualize the data using separate bar charts arranged horizontally (Fig. 1.5), with each chart representing the price of an individual stock. While this design effectively highlights the specific value of each stock, it struggles to convey the total market capitalization of each sector or industry. The total capitalization is represented by the combined area of all the horizontal bars. However, this design gave us the insight that using area as a key factor can effectively illustrate relative quantities in a visually intuitive and compelling way.

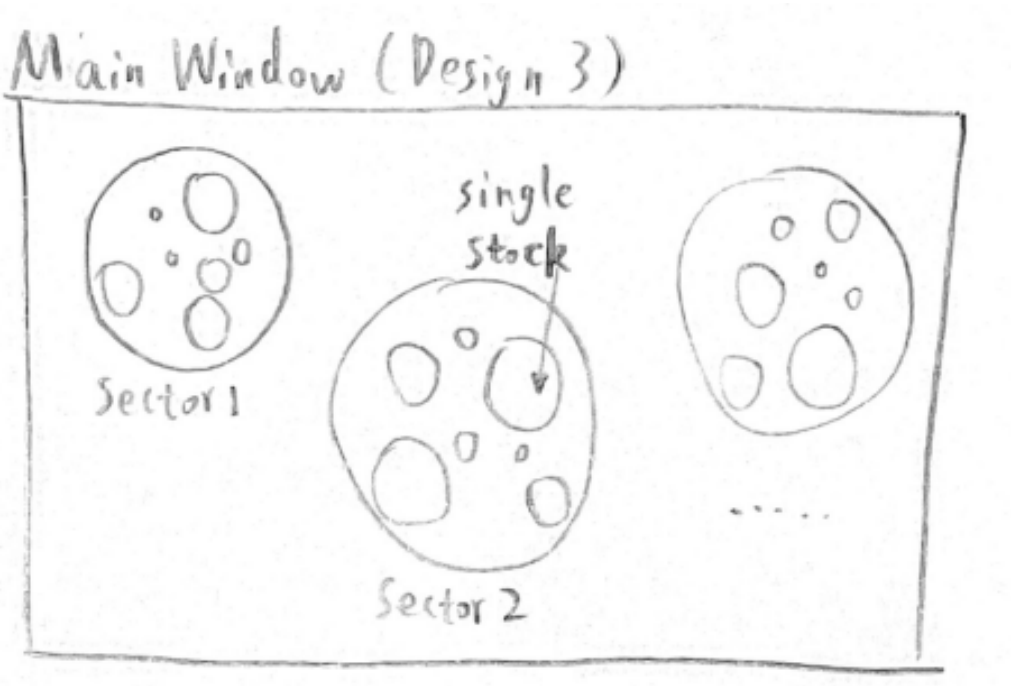


Figure 1.6: Design final: Using bubble diagram.

The final design involves using a clustered circle visualization to represent each sector's or industry's total capitalization, with smaller circles within each cluster representing individual stock prices (see Fig. 1.6). We believe this design offers two main advantages: First, it efficiently displays the relative value of each sector, as sectors or industries are naturally grouped together. Second, it clearly distinguishes each individual stock within the cluster. Additionally, by using circles to represent data, we eliminate the issue associated with using a vertical axis with a unified scale. Smaller stocks appear as points on the display, while more developed stocks are represented by larger circles. This dynamic representation is also intuitive as the user adjusts the timeline(see Fig. 1.7).

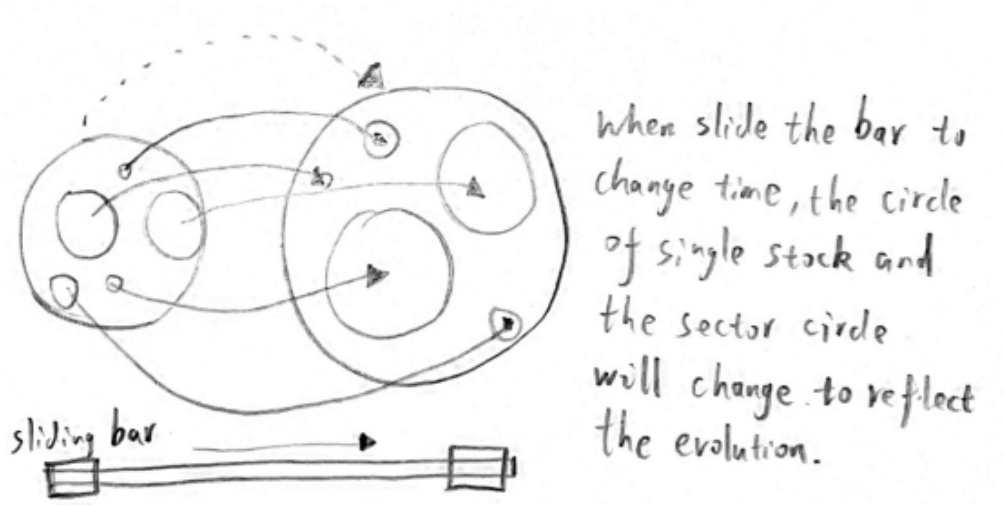


Figure 1.7: Dynamics of bubble diagram when sliding the time bar.

We also provided an stretch up of our optional feature (see Fig. 1.8): as the time bar is adjusted, a separate window will display the total capitalization value of each sector or industry using a line chart.

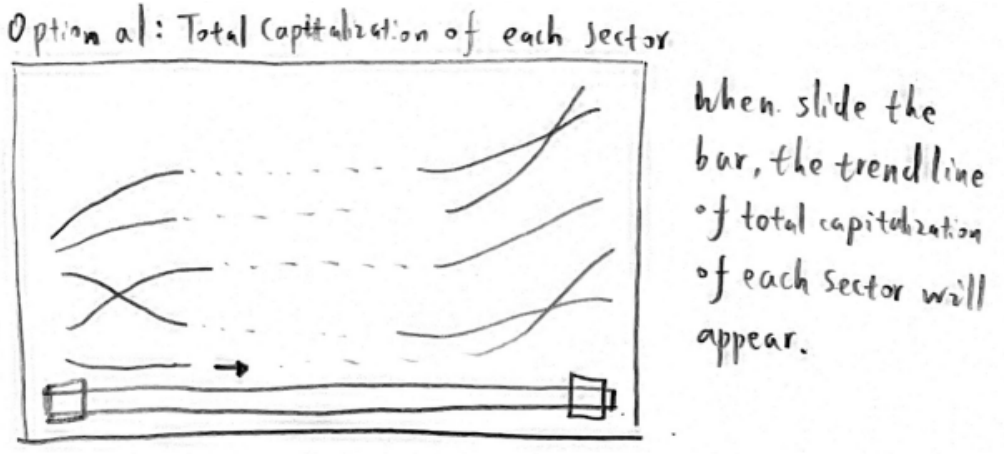


Figure 1.8: Optional feature: display the total capitalization.

1.7 Must-Have Features

Here we list some of the must have features:

1. At the bottom of the visualization windows, there should be a sliding timeline, which allows users to interact with the windows to show different time periods of the market.

2. The main window can display clustered data, each cluster is based on GICS sector/industry classifications.

3. In each cluster, there should be a set of circles to represent a single stock. The diameter of the circles should reflect the market capitalization of such a stock.

1.8 Option Features

Here we list some optional features.

1. Another separate window displays the total market capitalization of each sector and industry (summation of a single stock's capitalization within it), reflecting the overall economic development situation.

2. Another separate window displays the percentage of each sector or industry relative to the total market capitalization (summation of all).

1.9 Project Schedule

Our schedule:

1. Week 5-6: Finish the data collection and data processing.

2. Week 7-8: Add a sliding timeline, which allow the user to interact with. Main window can reflect the action of such time.

3. Week 9: Adding circles to display the stock data.

4. Week 10: Finalize the project to make it beautiful.

Project Review Feedback

This chapter is filled on Sep 27th, 2024.

2.1 Project Review Feedback

On Sep 27th, we scheduled a project review with the TA. During the meeting, we briefly introduced the motivation, functionality, and questions we were looking to answer. We then presented the interactive features we wanted to implement and the way the main interface would display the data. Lastly, we introduce the data we have collected so far and how we processed it to form a JSON storage structure that could be used in JavaScript. The TA acknowledged our progress and approved our project topic. He also provided several suggestions:

1. On the main interface, consider adding an introduction to the industrial classification standard, GICS.

2. When displaying stock market values at each time point, as presented in the initial plan, the main interface currently shows the market value of individual stocks in a scatter plot, but it lacks detailed information about individual stocks. If users are interested in more details, it is difficult to follow up. Therefore, additional stock information under the industry classification (such as company name, stock price, market value, trading liquidity, etc.) could be displayed in some side windows, or dynamic labels could be added to the main interface for further details.

Midterm Milestone

This chapter is filled on Oct 25th, 2024.

3.1 Current progress

At this midterm milestone, we have successfully launched a web-based visualization platform that effectively organizes and integrates our comprehensive dataset. The platform features a robust data storage architecture that seamlessly feeds into various visualization components.

The interactive interface we have developed demonstrates successful data binding and user engagement capabilities. Our primary visualization component incorporates a dynamic time slider, enabling users to explore market capitalization trends across different GICS industries at any given point in time.

Beyond the main "Overview" visualization, we have implemented detailed analytical views for individual GICS industries. These industry-specific visualizations consist of:

1. A bar chart displaying comparative market capitalization data
2. A line chart tracking historical performance trends
3. Detailed stock information for deeper analysis

This integrated approach provides users with both broad market perspectives and granular industry-specific insights, all accessible through an intuitive interface.

3.1.1 Functionality

There is a sliding bar on the bottom of the visualization window. The user can use this bar to adjust time to view dynamic process.

3.1.2 Overview

The overview visualization takes all the single stock within the GICS industry as a cluster with each circle's radius reflects the capitalization of the single stock.

3.1.3 GICS Single Industry and Stock Details

When the user click the side contents of the GICS industry, the main visualization will display the details of single stock within this industry. The bottom view will also show history information of the stocks.

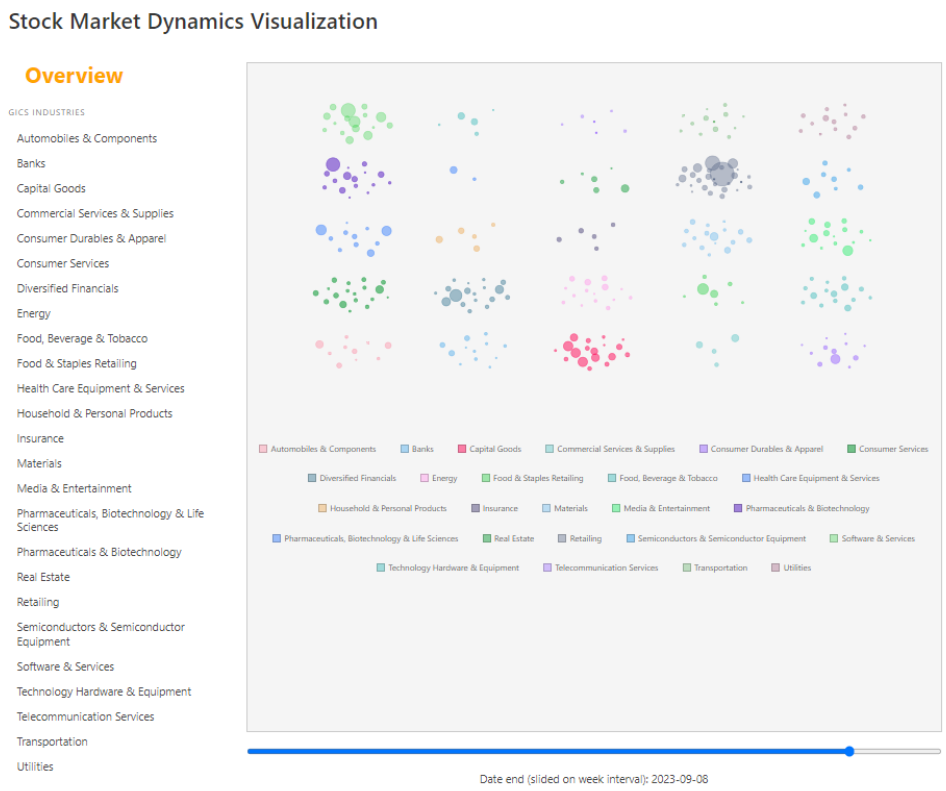
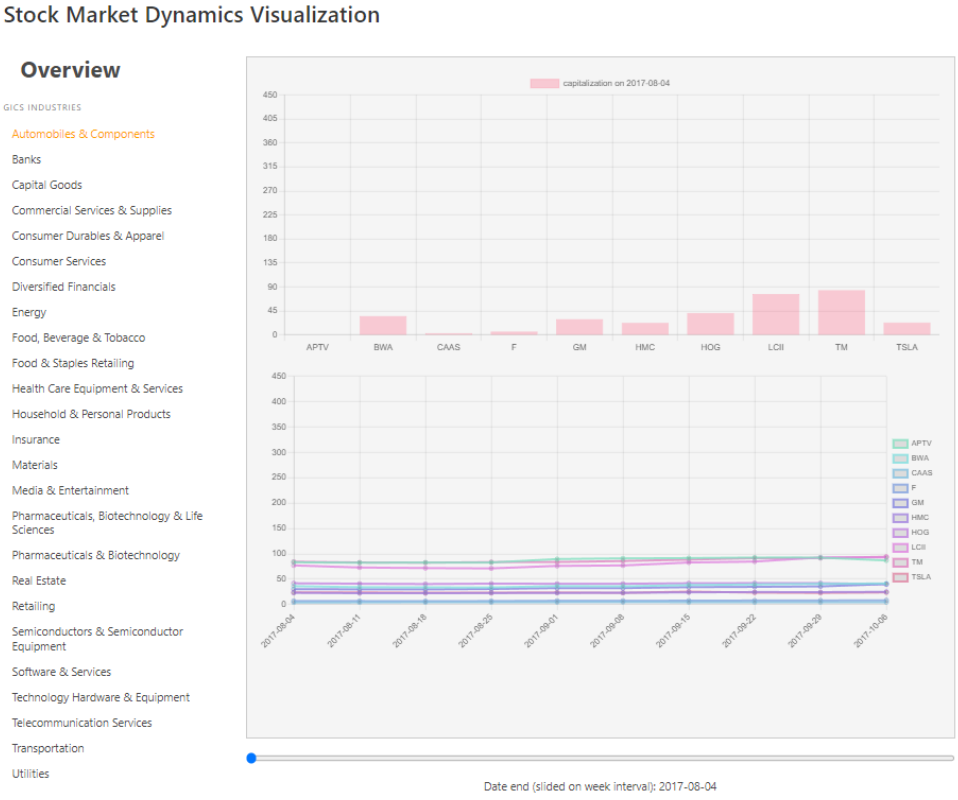
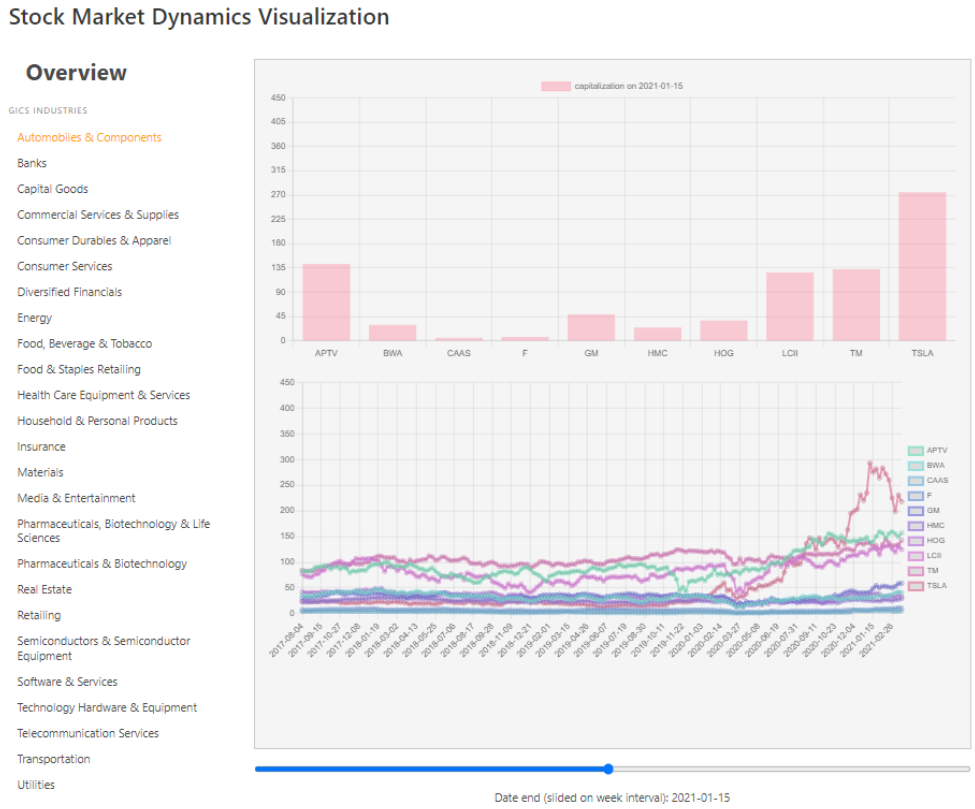
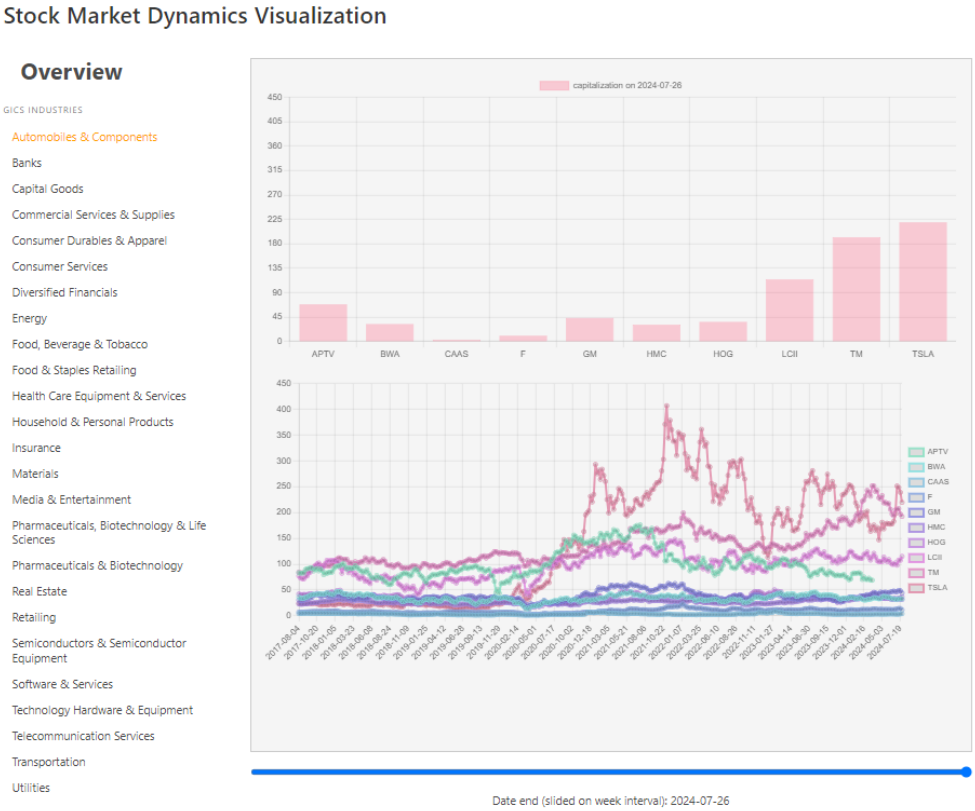


Figure 3.1: Overview: this visualization can display the result as an overview to each of the GICS industry.




Stock Market Dynamics Visualization
Overview
GICS INDUSTRIES
Automobiles & Components
Banks
Capital Goods
Commercial Services & Supplies
Consumer Durables & Apparel
Consumer Services
Diversified Financials
Energy
Food, Beverage & Tobacco
Food & Staples Retailing
Health Care Equipment & Services
Household & Personal Products
Insurance
Materials
Media & Entertainment
Pharmaceuticals, Biotechnology & Life Sciences
Pharmaceuticals & Biotechnology
Real Estate
Retailing
Semiconductors & Semiconductor Equipment
Software & Services
Technology Hardware & Equipment
Telecommunication Services
Transportation
Utilities
capitalization on 2021-01-15
APTV
BWA
CAAS
F
GM
HMC
HOG
LCII
TM
TSLA
Date end (slided on week interval): 2021-01-15


Stock Market Dynamics Visualization
Overview
GICS INDUSTRIES
Automobiles & Components
Banks
Capital Goods
Commercial Services & Supplies
Consumer Durables & Apparel
Consumer Services
Diversified Financials
Energy
Food, Beverage & Tobacco
Food & Staples Retailing
Health Care Equipment & Services
Household & Personal Products
Insurance
Materials
Media & Entertainment
Pharmaceuticals, Biotechnology & Life Sciences
Pharmaceuticals & Biotechnology
Real Estate
Retailing
Semiconductors & Semiconductor Equipment
Software & Services
Technology Hardware & Equipment
Telecommunication Services
Transportation
Utilities
capitalization on 2024-07-26
APTV
BWA
CAAS
F
GM
HMC
HOG
LCII
TM
TSLA
Date end (slided on week interval): 2024-07-26

3.1.4 Future Plan

The current release serves as an interim milestone, with several key enhancements planned for the next version:

1. **Enhanced Layout Organization.** We plan to restructure the page layout to achieve better integration between GICS industry categories, stock details, and the overview visualization. This reorganization aims to provide a more intuitive and cohesive user experience.

2. **Advanced Overview Visualization.** The current overview display shows individual stock bubbles based on market capitalization. In the next iteration, we will implement hierarchical clustering, where each GICS industry will be represented by a larger encompassing bubble, with its radius proportional to the industry's total market capitalization, while maintaining individual stock representations within each cluster.

3. **Comprehensive Axis Labeling.** We will enhance the chart components by ensuring all axes are properly labeled and visible, particularly focusing on adding clear y-axis labels and scales to improve data interpretation and readability.

Overview and Motivation

This chapter contains similar content in project proposal. Please also view the proposal.

4.1 Overview and Motivation

The stock market serves as a vital indicator of economic health and industrial development trends. Our project aims to create an interactive visualization system that helps investors and analysts understand the dynamic evolution of different industries through the lens of market capitalization. By visualizing how various sectors of the economy grow, shrink, and interact over time, we provide insights into the broader patterns of economic development and industrial transformation.

The motivation for this project stems from the challenge investors face in understanding macro-level industrial shifts. Over time spans ranging from several years to decades, industries emerge, flourish, decline, or transform. Traditional methods of analyzing these changes often involve examining individual stocks or basic sector indices, which can miss the broader patterns of industrial evolution. Our visualization system addresses this gap by providing an intuitive, interactive way to observe how different sectors of the economy develop over time.

We specifically chose to focus on market capitalization as our primary metric because it effectively reflects both the current value and future expectations of different industries. By organizing this data according to the Global Industry Classification Standard (GICS), we provide a standardized framework that allows users to track and compare industrial development across different sectors of the economy.

4.2 Related Work

Our visualization approach draws inspiration from several key sources in both academic literature and practical applications:

Treemaps in Financial Visualization: The SmartMoney Map of the Market, introduced in the late 1990s, pioneered the use of treemaps to display market capitalization data. Their approach to showing hierarchical financial data influenced our decision to use nested visualizations for industry groupings.

Force-Directed Layouts: Our clustering approach for the overview visualization was inspired by force-directed graph layouts used in network visualization. Similar techniques have been employed by financial data providers like Bloomberg and Reuters to show market relationships.

Academic Visualization Research: The bubble chart layout we implemented draws from research on space-efficient visualization techniques, particularly the work on "Bubble Trees" for hierarchical data visualization. This influenced our decision to use nested circles to represent market capitalization at both industry and individual stock levels.

Industry Standards: The GICS classification system, developed by MSCI and S&P Global, provides the foundational structure for our visualization. This widely-adopted standard

ensures our visualization aligns with industry practices and professional analysis frameworks.

4.3 Questions

Our project aimed to address several key questions about industrial dynamics and market evolution:

Primary Questions:

- How do different industries evolve over time in terms of market capitalization?
- What patterns emerge in the relative growth and decline of different sectors?
- How does the composition of major industries change over time?
- What is the relationship between individual company growth and overall industry development?

As the project evolved, new questions emerged that expanded our investigation:

Emerging Questions:

- How do economic cycles affect different industries differently?
- What role do dominant companies play in shaping their industry's overall market capitalization?
- How does the concentration of market capitalization within industries change over time?

These questions shaped our visualization design choices. For example, the need to show both individual company performance and industry-wide trends led to our dual-view approach with both overview and detailed visualizations. The temporal aspect of our questions influenced our decision to implement an interactive timeline that allows users to explore changes over time.

The exploratory nature of these questions also influenced our decision to implement interactive features that allow users to investigate different aspects of the data. The ability to switch between industry views and examine specific time periods enables users to pursue their own questions and discover patterns that might not be immediately apparent in static visualizations.

Data Processing

This chapter is filled on Oct.19, 2024.

5.1 The meta-data required

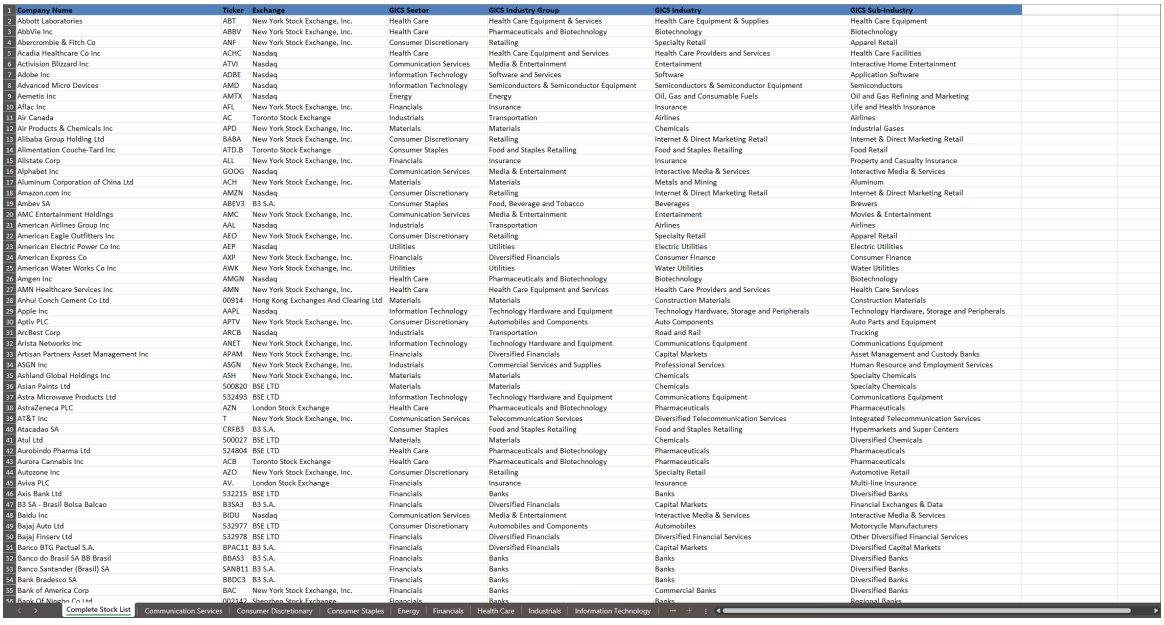
In the previous section of the project proposal, we confirmed that this project are going to use the U.S. stock market capitalization to reflect industry development trends and capital flow (using GICS industry classification standard). Therefore, we first need to collect and manipulate the data.

Specifically, the following data is required:

1. GICS industry classification stock codes
2. The meta-data of all stocks through these codes
3. Stock prices calculated from the meta-data
4. (Optional) Other clustering data, such as the total value of a specific industry

5.2 The Crawler and feeding data structure

Firstly, we researched GICS industry classification data and used data from here to obtain the stock codes classified by GICS sector and industry categories within the S&P 500 index. At the same time, we made data cleaning to narrow to the U.S. market stocks.



Company Name	Ticker	Exchange	GICS Sector	GICS Industry Group	GICS Industry	GICS Sub-Industry
Abbott Laboratories	ABT	New York Stock Exchange, Inc.	Health Care	Health Care Equipment & Services	Health Care Equipment & Supplies	Health Care Equipment
AbbVie Inc	ABBV	New York Stock Exchange, Inc.	Health Care	Pharmaceuticals and Biotechnology	Biotechnology	Biotechnology
Abercrombie & Fitch Co	ANF	New York Stock Exchange, Inc.	Consumer Discretionary	Retailing	Specialty Retail	Apparel Retail
Acadia Healthcare Co Inc	ACHC	Nasdaq	Health Care	Health Care Equipment and Services	Health Care Providers and Services	Health Care Facilities
Activision Blizzard Inc	ATVI	Nasdaq	Communication Services	Media & Entertainment	Entertainment	Interactive Home Entertainment
Adobe Inc	ADBE	Nasdaq	Information Technology	Software and Services	Software	Application Software
Advanced Micro Devices	AMD	Nasdaq	Information Technology	Semiconductors & Semiconductor Equipment	Semiconductors & Semiconductor Equipment	Semiconductors
Aemetis Inc	AMTX	Nasdaq	Energy	Energy	Oil, Gas and Consumable Fuels	Oil and Gas Refining and Marketing
Aflac Inc	AFL	New York Stock Exchange, Inc.	Financials	Insurance	Insurance	Life and Health Insurance
Air Canada	AC	Toronto Stock Exchange	Industrials	Transportation	Airlines	Airlines
Air Products & Chemicals Inc	APD	New York Stock Exchange, Inc.	Materials	Materials	Chemicals	Industrial Gases
Alibaba Group Holding Ltd	BABA	New York Stock Exchange, Inc.	Consumer Discretionary	Retailing	Internet & Direct Marketing Retail	Internet & Direct Marketing Retail
Alimentation Couche-Tard Inc	ATD.B	Toronto Stock Exchange	Consumer Staples	Food and Staples Retailing	Food and Staples Retailing	Food Retail
Allstate Corp	ALL	New York Stock Exchange, Inc.	Financials	Insurance	Insurance	Property and Casualty Insurance
Alphabet Inc	GOOG	Nasdaq	Communication Services	Media & Entertainment	Interactive Media & Services	Interactive Media & Services
Aluminum Corporation of China Ltd	ACH	New York Stock Exchange, Inc.	Materials	Materials	Metals and Mining	Aluminum
Amazon.com Inc	AMZN	Nasdaq	Consumer Discretionary	Retailing	Internet & Direct Marketing Retail	Internet & Direct Marketing Retail
Ambev SA	ABEV3	B3 S.A.	Consumer Staples	Food, Beverage and Tobacco	Beverages	Brewers
AMC Entertainment Holdings	AMC	New York Stock Exchange, Inc.	Communication Services	Media & Entertainment	Entertainment	Movies & Entertainment
American Airlines Group Inc	AAL	Nasdaq	Industrials	Transportation	Airlines	Airlines
American Eagle Outfitters Inc	AEO	New York Stock Exchange, Inc.	Consumer Discretionary	Retailing	Specialty Retail	Apparel Retail
American Electric Power Co Inc	AEP	Nasdaq	Utilities	Utilities	Electric Utilities	Electric Utilities
American Express Co	AXP	New York Stock Exchange, Inc.	Financials	Diversified Financials	Consumer Finance	Consumer Finance
American Water Works Co Inc	AWK	New York Stock Exchange, Inc.	Utilities	Utilities	Water Utilities	Water Utilities
Amgen Inc	AMGN	Nasdaq	Health Care	Pharmaceuticals and Biotechnology	Biotechnology	Biotechnology
AMN Healthcare Services Inc	AMN	New York Stock Exchange, Inc.	Health Care	Health Care Equipment and Services	Health Care Providers and Services	Health Care Services
Anhui Conch Cement Co Ltd	00914	Hong Kong Exchanges And Clearing Ltd	Materials	Materials	Construction Materials	Construction Materials
Apple Inc	AAPL	Nasdaq	Information Technology	Technology Hardware and Equipment	Technology Hardware, Storage and Peripherals	Technology Hardware, Storage and Peripherals
Aptiv PLC	APTV	New York Stock Exchange, Inc.	Consumer Discretionary	Automobiles and Components	Auto Components	Auto Parts and Equipment
ArcBest Corp	ARCB	Nasdaq	Industrials	Transportation	Road and Rail	Trucking
Arista Networks Inc	ANET	New York Stock Exchange, Inc.	Information Technology	Technology Hardware and Equipment	Communications Equipment	Communications Equipment
Artisan Partners Asset Management Inc	APAM	New York Stock Exchange, Inc.	Financials	Diversified Financials	Capital Markets	Asset Management and Custody Banks
ASGN Inc	ASGN	New York Stock Exchange, Inc.	Industrials	Commercial Services and Supplies	Professional Services	Human Resource and Employment Services
Ashland Global Holdings Inc	ASH	New York Stock Exchange, Inc.	Materials	Materials	Chemicals	Specialty Chemicals
Asian Paints Ltd	500820	BSE LTD	Materials	Materials	Chemicals	Specialty Chemicals
Astra Microwave Products Ltd	532493	BSE LTD	Information Technology	Technology Hardware and Equipment	Communications Equipment	Communications Equipment
AstraZeneca PLC	AZN	London Stock Exchange	Health Care	Pharmaceuticals and Biotechnology	Pharmaceuticals	Pharmaceuticals
AT&T Inc	T	New York Stock Exchange, Inc.	Communication Services	Telecommunication Services	Diversified Telecommunication Services	Integrated Telecommunication Services
Atacadao SA	CRFB3	B3 S.A.	Consumer Staples	Food and Staples Retailing	Food and Staples Retailing	Hypermarkets and Super Centers
Atul Ltd	500027	BSE LTD	Materials	Materials	Chemicals	Diversified Chemicals
Aurobindo Pharma Ltd	524804	BSE LTD	Health Care	Pharmaceuticals and Biotechnology	Pharmaceuticals	Pharmaceuticals
Aurora Cannabis Inc	ACB	Toronto Stock Exchange	Health Care	Pharmaceuticals and Biotechnology	Pharmaceuticals	Pharmaceuticals
Autozone Inc	AZO	New York Stock Exchange, Inc.	Consumer Discretionary	Retailing	Specialty Retail	Automotive Retail
Aviva PLC	AV.	London Stock Exchange	Financials	Insurance	Insurance	Multi-line Insurance
Axis Bank Ltd	532215	BSE LTD	Financials	Banks	Banks	Diversified Banks
B3 SA - Brasil Bolsa Balcao	B3SA3	B3 S.A.	Financials	Diversified Financials	Capital Markets	Financial Exchanges & Data
Baidu Inc	BIDU	Nasdaq	Communication Services	Media & Entertainment	Interactive Media & Services	Interactive Media & Services
Bajaj Auto Ltd	532977	BSE LTD	Consumer Discretionary	Automobiles and Components	Automobiles	Motorcycle Manufacturers
Bajaj Finserv Ltd	532978	BSE LTD	Financials	Diversified Financials	Diversified Financial Services	Other Diversified Financial Services
Banco BTG Pactual S.A.	BPAC11	B3 S.A.	Financials	Diversified Financials	Capital Markets	Diversified Capital Markets
Banco do Brasil SA BB Brasil	BBAS3	B3 S.A.	Financials	Banks	Banks	Diversified Banks
Banco Santander (Brasil) SA	SANB11	B3 S.A.	Financials	Banks	Banks	Diversified Banks
Bank Bradesco SA	BBDC3	B3 S.A.	Financials	Banks	Banks	Diversified Banks
Bank of America Corp	BAC	New York Stock Exchange, Inc.	Financials	Banks	Commercial Banks	Diversified Banks
Bank Of Ningbo Co Ltd	002142	Shenzhen Stock Exchange	Financials	Banks	Banks	Regional Banks

Complete Stock List | Communication Services | Consumer Discretionary | Consumer Staples | Energy | Financials | Health Care | Industrials | Information Technology

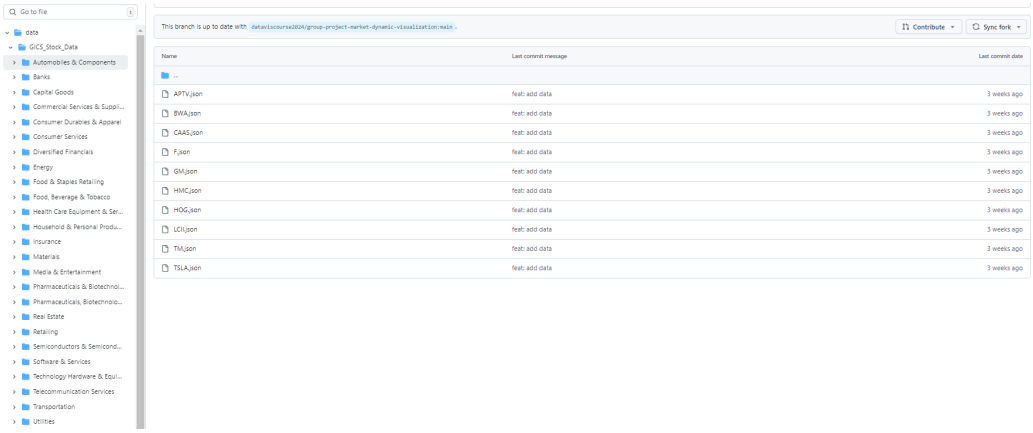


Figure 5.1: The data storage structure.


```
[
    {
        "date":"2017-12-08",
        "open":102.39,
        "close":83.24,
        "high":102.39,
        "low":81.89,
        "volume":11854567,
        "turnover":1023377226.0,
        "amplitude":136.48,
        "price_change_percent":454.19,
        "price_change_amount":68.22,
        "turnover_rate":4.46
    },
    {
        "date":"2017-12-15",
        "open":84.32,
        "close":82.42,
        "high":84.96,
        "low":80.99,
        "volume":13448660,
        "turnover":1142879988.0,
        "amplitude":4.77,
        "price_change_percent":-0.99,
        "price_change_amount":-0.82,
        "turnover_rate":5.06
    },
    {
        "date":"2017-12-22",
        "open":82.6,
        "close":82.81,
        "high":84.93,
        "low":82.15,
        "volume":8125324,
        "turnover":694842682.0,
        "amplitude":3.37,
        "price_change_percent":0.47,
        "price_change_amount":0.39,
        "turnover_rate":3.06
```


Figure 5.2: The meta-data for each stock.

Further, we used the AKShare API to collect the interval data of these stocks from 2017 to 2024. We stored these meta-data in a JSON file based on GICS industry classification and processed them with database scripts (e.g., JavaScript) for our webpage usage.

Exploratory Data Analysis and Design Evolution

6.1 Exploratory Data Analysis

6.1.1 Initial Exploratory Data Analysis

Our exploratory data analysis began by examining individual stock performance before progressing to more complex industry-wide visualizations:

1. **Individual Stock Analysis:** Our first visualization focused on understanding single stock behavior through two key metrics:

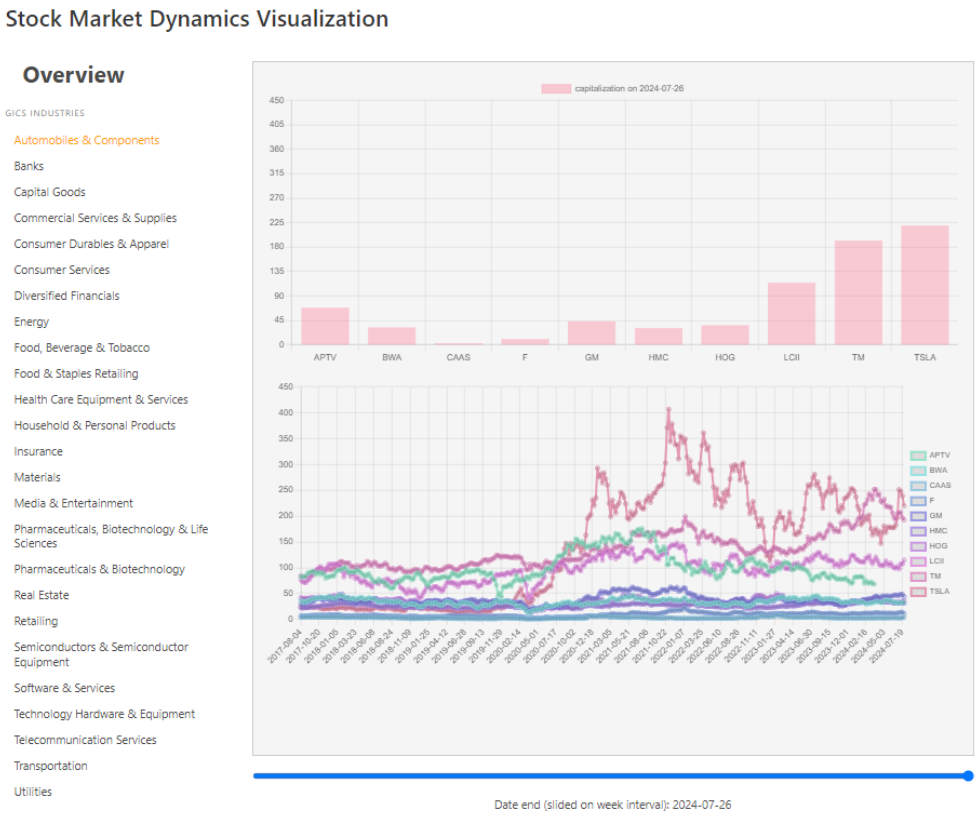


Figure 6.1: Individual stock market capitalization and price visualization

We implemented a dual-chart system showing:

- Bar chart displaying current market capitalization
- Line chart tracking historical stock price movements

This approach provided clear insights into individual stock performance but didn't address industry-level patterns.

2. **Initial Bubble Chart:** We then developed a bubble visualization to show multiple stocks simultaneously:

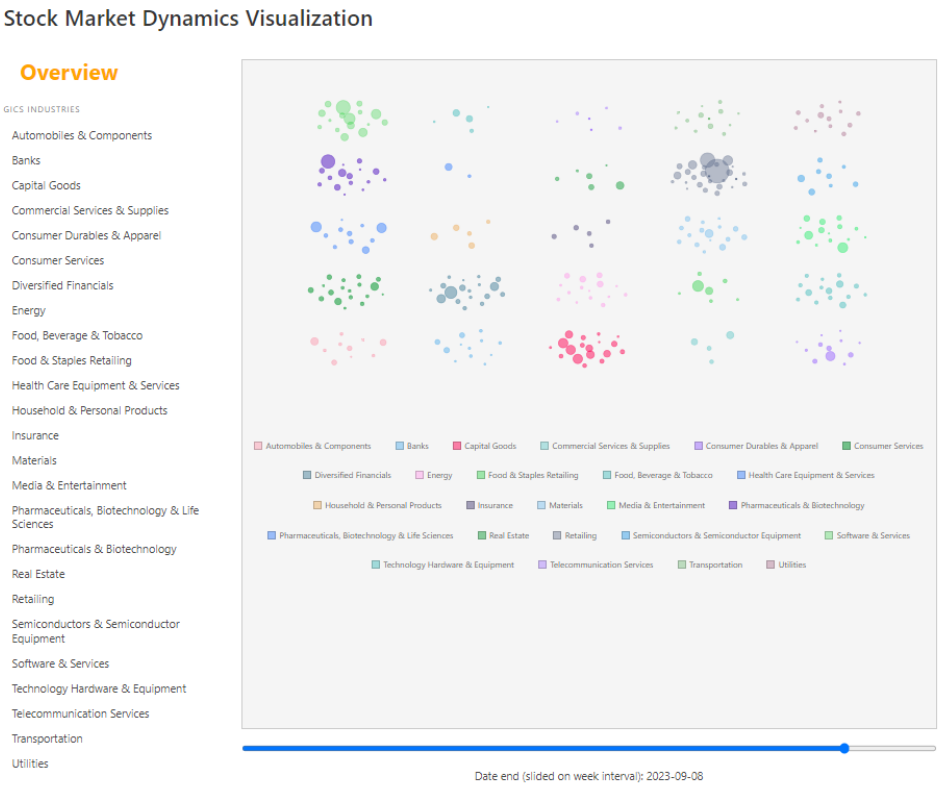


Figure 6.2: Initial bubble chart showing individual stocks

This version:

- Represented each stock as a bubble sized by market capitalization
- Grouped stocks by GICS industry
- Used color coding to distinguish different industries

However, this visualization lacked the ability to show industry-level aggregations. So on the next step, we will add total market capitalization of the industrial group to showcase the aggregated results.

This is the initial data analysis approach we implemented. This progression from individual stock analysis to complex multi-level visualization helped us understand both the detailed behavior of individual stocks and their relationships within their respective industries. Each iteration revealed new insights and challenges that informed our subsequent design decisions.

6.1.2 Key Insights

Through our exploratory analysis, we gained several crucial insights:

(a) **Scale Disparities:** We discovered enormous variations in market capitalization across different industries and companies. For example, some technology companies had market caps orders of magnitude larger than entire traditional industries.

(b) **Temporal Patterns:** Our time-series analysis revealed distinct growth patterns:

- Technology and healthcare sectors showed consistently high growth rates
- Traditional industries exhibited more stable, gradual changes
- Certain sectors showed clear cyclical patterns

(c) **Industry Clustering:** We observed natural groupings within industries:

- Most industries had a few dominant large-cap companies
- Mid-cap companies often showed similar behavior within sectors
- Small-cap companies tended to have higher volatility

6.1.3 Design Implications

These insights directly influenced our final design decisions:

(a) **Hierarchical Visualization:** The wide range of market capitalizations led us to adopt a bubble chart approach that could effectively show both large and small values simultaneously.

(b) **Interactive Timeline:** Observing the rich temporal patterns in our data, we implemented an interactive timeline to allow users to explore these changes over time.

(c) **Dual-View System:** We developed a two-part visualization system:

- Overview: Showing industry-level patterns and relationships
- Detail View: Displaying individual stock performance within each industry

(d) **Color Coding:** We implemented a consistent color scheme for industries to help users track patterns across different views and time periods.

The exploratory process led us to develop a visualization system that effectively balances the needs to show both macro-level industry trends and micro-level individual stock performance. The final design allows users to explore the complex relationships between industries while maintaining the ability to drill down into specific companies of interest.

6.2 Design Evolution

Our design began with several visualization approaches to understand the complex relationships between industries and their constituent stocks. We experimented with three primary visualizations:

(a) **Stacked Bar Charts:** Our first attempt used traditional stacked bar charts to represent market capitalization across industries. Each bar represented an industry, with segments showing individual stock contributions:

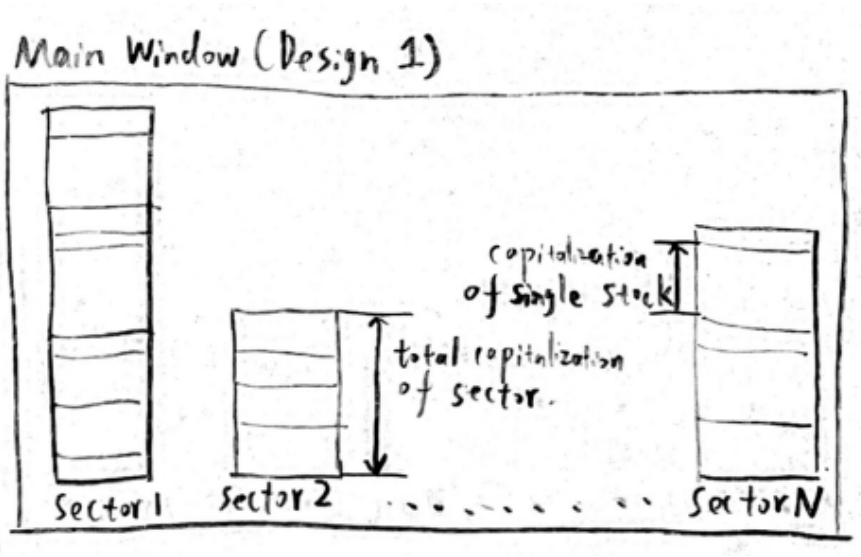


Figure 6.3: Stacked bar visualization draft

While this approach clearly showed the total market capitalization of each industry, we discovered two significant limitations:

- When certain companies (like major tech firms) grew rapidly, the scale adjustments made smaller companies virtually invisible
- The stacking made it difficult to compare individual stocks across different industries

(b) **Separated Bar Charts:** To address these limitations, we explored using separate bars for each stock:

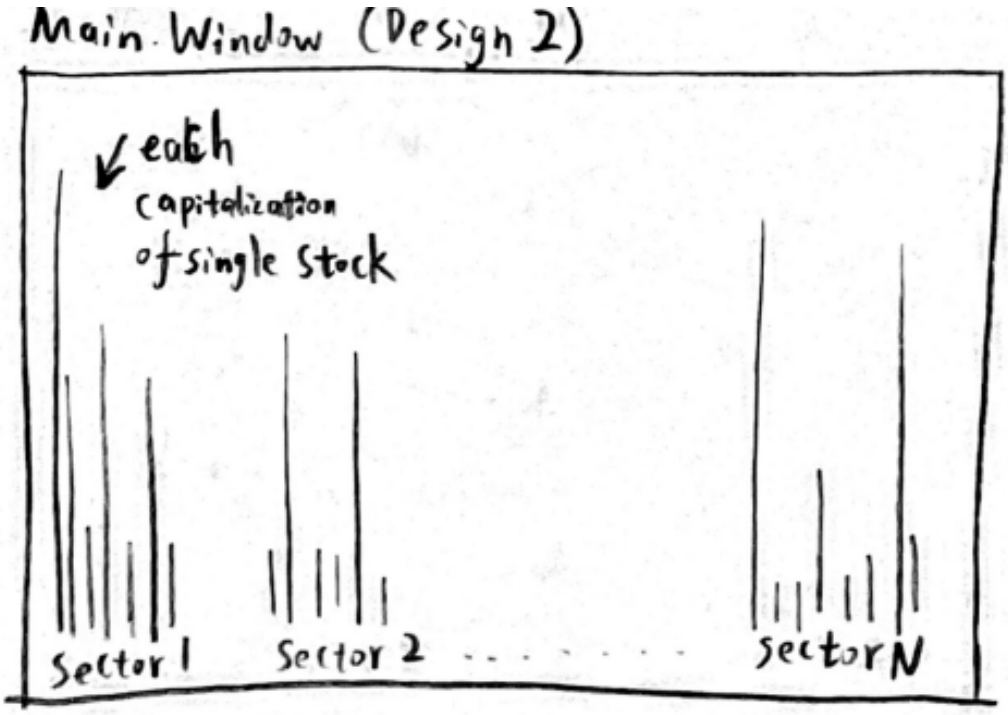


Figure 6.4: Separated bar chart design draft

This improved individual stock visibility but created new challenges:

- Difficult to grasp total industry size
- Space inefficiency with many stocks
- Limited ability to show temporal changes

(c) **Bubble Charts:** Finally, we experimented with bubble charts for their ability to represent hierarchical data:

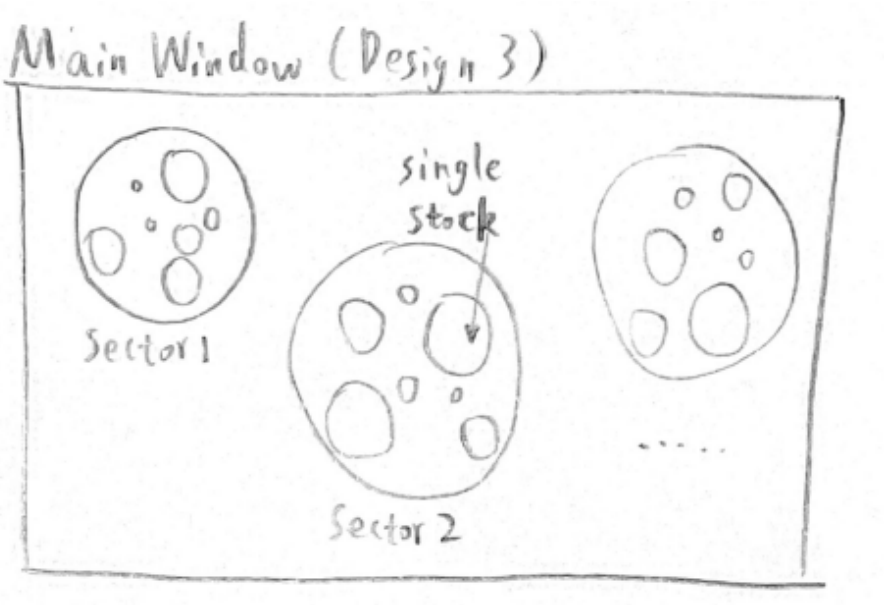


Figure 6.5: Final bubble chart visualization design

Implementation

7.1 Features

This visualization tool is designed to provide intuitive insights into stock market dynamics across different GICS sectors. The application consists of three main pages: Introduction, Overview, and Subsector pages, each serving distinct purposes in data presentation and analysis. Also, we embedded a video of screen to introduce our features.

7.1.1 Introduction Page

The Introduction page serves as an entry point for users to understand the GICS (Global Industry Classification Standard) framework and its significance in market analysis.

Components

- **GICS Explanation**: A comprehensive description of the GICS framework and its hierarchical structure with figures.
- **Demo Video**: Embedded YouTube video providing project demo.

7.1.2 Overview Page

The Overview page provides a high-level visualization of market dynamics across all GICS sectors.

Components

- **Bubble Chart**: Interactive visualization showing market capitalization and sector relationships
- **Time Slider**: Dynamic control for temporal analysis of market data
- **Color-coded Legend**: Sector identification through consistent color schemes

7.1.3 Subsector Pages

Individual sector pages offering detailed analysis of specific GICS industries.

Components

- **Bar Chart**: Company-wise market capitalization comparison
- **Line Chart**: Historical price trends for selected companies
- **Interactive Elements**: Tooltips and clickable elements for detailed information
- **Synchronized Updates**: Coordinated visualization updates across charts

7.2 Pipeline and Data Flow

Our visualization system implements a data pipeline that transforms raw stock market data into interactive visualizations. The pipeline consists of three main stages: data acquisition, processing, and visualization rendering.

At the data acquisition stage, we first collect stock codes based on the GICS (Global Industry Classification Standard) classification system. The stock data is organized hierarchically, with each stock assigned to specific industry categories. Our system queries financial data through a combination of static JSON files and dynamic data loading:

```
fetch('data/filtered_stock_list.json')
    .then(response => response.json())
    .then(data => {
        companyCodeToName = data;
    })
    .catch(error => console.error(
    'Error loading company code to name data:', error));
```

The data processing stage involves several transformations to prepare the raw data for visualization. For each industry, we process stock data to calculate key metrics such as market capitalization. The calculation is performed using a standardized formula:

```
function calculateMarketCap(turnover, turnoverRate, price) {
    if (!turnoverRate || turnoverRate === 0) return 0;
    const marketCap = (turnover / turnoverRate) * price;
    return marketCap / 1000000000; // Convert to billions
}
```

The data flow follows a hierarchical structure in which industry-level data is composed of aggregated company-level data. When users interact with the visualization, the data flows through several stages.

(a) Initial stock list loading and GICS classification mapping

(b) Individual stock data retrieval from JSON files based on industry selection

(c) Market capitalization calculation and data transformation

(d) Visualization rendering with appropriate scales and layouts

For the overview visualization, we implement a specialized data flow that handles multiple industries simultaneously:

```
async function initializeCache() {
    const industryNames = Object.keys(gicsIndustryData);
    for (const industryName of industryNames) {
        const companies = gicsIndustryData[industryName];
        for (const company of companies) {
            const stockFilePath =
                `data/GICS_Stock_Data/${industryName}/${company}.json`;
            try {
                const response = await fetch(stockFilePath);
                const data = await response.json();
                cachedStockData[company] = {
                    data: data,
                    industry: industryName,
                    color: industryColors[industryName]
                };
            } catch (error) {
                console.error(
                `Error loading data for ${company}:`, error);
                cachedStockData[company] = null;
            }
        }
    }
}
```

The visualization rendering stage employs different chart types depending on the context. For the overview, we use a bubble chart layout that positions companies within their industry clusters. For detailed views, we implement both bar charts for current market capitalization and line charts for historical trends. The data flow is optimized through caching mechanisms that store frequently accessed data and pre-computed visualization parameters.

This pipeline architecture ensures efficient data handling while maintaining responsiveness during user interactions. The modular design allows for easy extension and modification of individual components without affecting the overall system integrity.

7.3 Implementation Details

7.3.1 Optimization

In our project's implementation, we focused on optimizing the data loading and visualization performance to ensure smooth user interactions, particularly when handling large datasets across multiple GICS industries. A key optimization strategy we employed was implementing a comprehensive caching for both the raw stock data and computed visualizations.

The caching mechanism operates on multiple levels. At the data level, we implemented the `cachedIndustryData` and `cachedStockData` objects to store previously loaded stock information. When users select different industries or time periods, the system first

checks these caches before making new data requests. This approach significantly reduces redundant data loading operations and improves response times, especially when users frequently switch between different views.

For the bubble chart visualization, we introduced the `cachedBubbleData` structure to store pre-computed bubble positions and sizes. This optimization was particularly crucial for the overview visualization, where we need to display multiple industries simultaneously. We also implemented the `bubblePositions` cache to maintain consistent positioning of industry clusters and individual stock bubbles across different time periods, ensuring smooth transitions and preventing disorienting layout shifts.

The optimization strategy is implemented through careful cache initialization:

```
if (Object.keys(bubblePositions).length === 0) {
    initializePositions();
}
if (Object.keys(cachedStockData).length === 0) {
    initializeCache().then(() => {
        const bubbleData = createBubbleDataFromCache(
            dates[timeline.value]
        );
        drawBubbleChart(bubbleData, bubbleCtx);
    });
}
```

7.3.2 Beautify

Our visualization interface prioritizes both aesthetic appeal and functional clarity through careful layout design and thoughtful visual hierarchy. The main interface is structured into three primary components: a navigation sidebar for industry selection, the main visualization area, and an interactive timeline slider at the bottom.

For the color scheme, we implemented a consistent industry-specific color coding system using the `industryColors` mapping. Each GICS industry is assigned a unique color that persists across different visualization components, helping users maintain context when analyzing different views. The colors are carefully chosen to be visually distinct while maintaining harmony across the interface.

Bubble visualization employs sophisticated layout algorithms to prevent overcrowding and ensure optimal use of space. We implemented an expansion factor system that dynamically adjusts bubble sizes and spacing.

```
const expansionFactor = 0.7;
const adjustedX = 50 + (centerX - 50) * expansionFactor;
const adjustedY = 50 + (centerY - 50) * expansionFactor;
```

Charts are enhanced with smooth animations and transitions, although these are carefully controlled to maintain performance:

```
options: {
    responsive: true,
    maintainAspectRatio: false,
```

```
    animation: {
        duration: 0
    }
}
```

For the industry-specific views, we implemented a dual-chart system showing both current market capitalization (bar chart) and historical trends (line chart). These charts are positioned in a vertically stacked layout with carefully calculated proportions to maximize readability while maintaining an aesthetically pleasing arrangement. The charts include interactive tooltips that provide detailed information on hover, enhancing the user experience without cluttering the visual space.

Final results and evaluations

8.1 Final Results

8.1.1 Visualization Components

Our final visualization system consists of three main components:

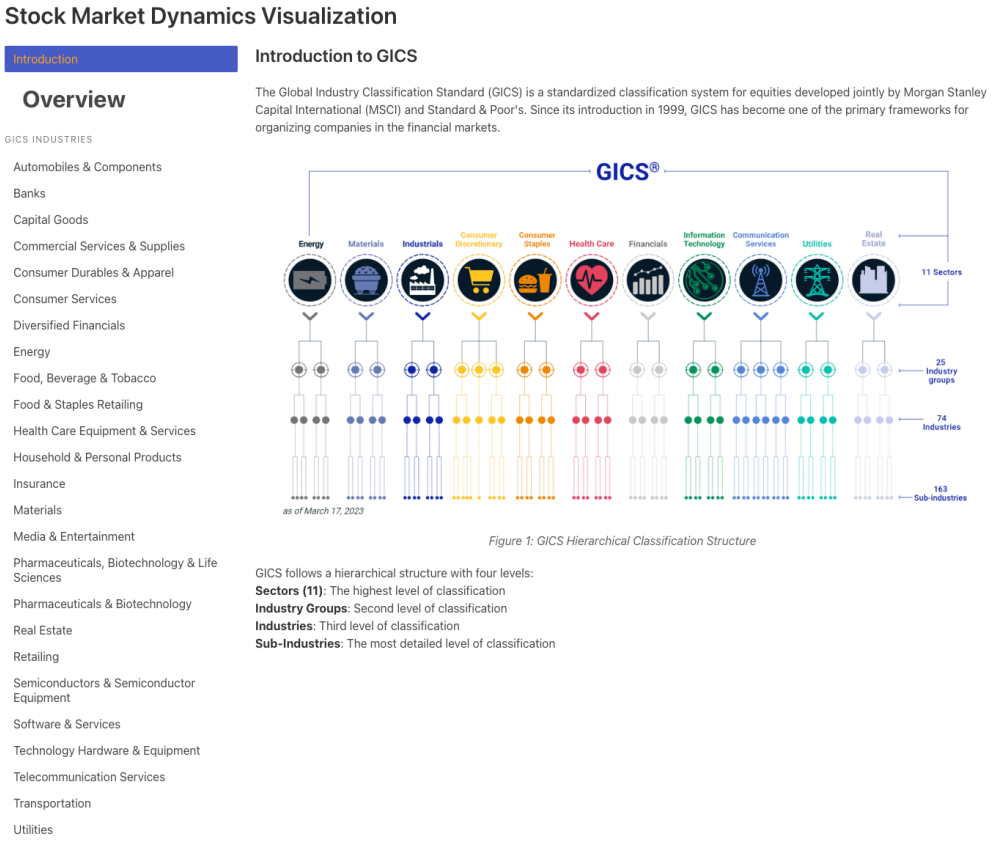


Figure 8.1: GICS Introduction and Classification System

The introduction page provides users with essential context about the GICS classification system, helping them understand how industries are organized and categorized in our visualization. This foundational knowledge is crucial for interpreting the subsequent visualizations.

Also, on the bottom of this page, we also provided a demo video to showcase the function of our site.

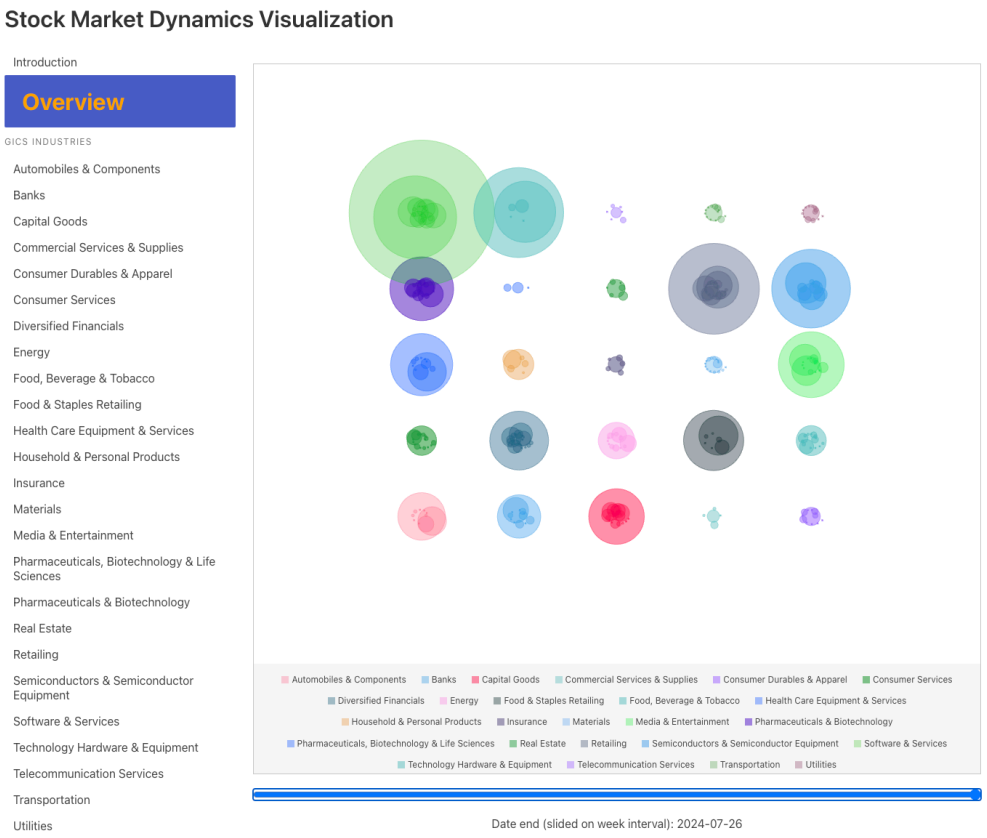


Figure 8.2: Market Overview Visualization

The overview visualization uses a bubble chart layout where each cluster represents a GICS industry, and individual bubbles within clusters represent specific companies. Through this visualization, we discovered several interesting patterns:

- Technology sector dominance, particularly evident in the large bubble sizes of companies like Microsoft (MSFT) and Apple (AAPL)
- Clear industry clustering patterns, with related sectors often showing similar growth trajectories
- Significant size disparities between different industries, with some traditional sectors showing relatively smaller market capitalizations

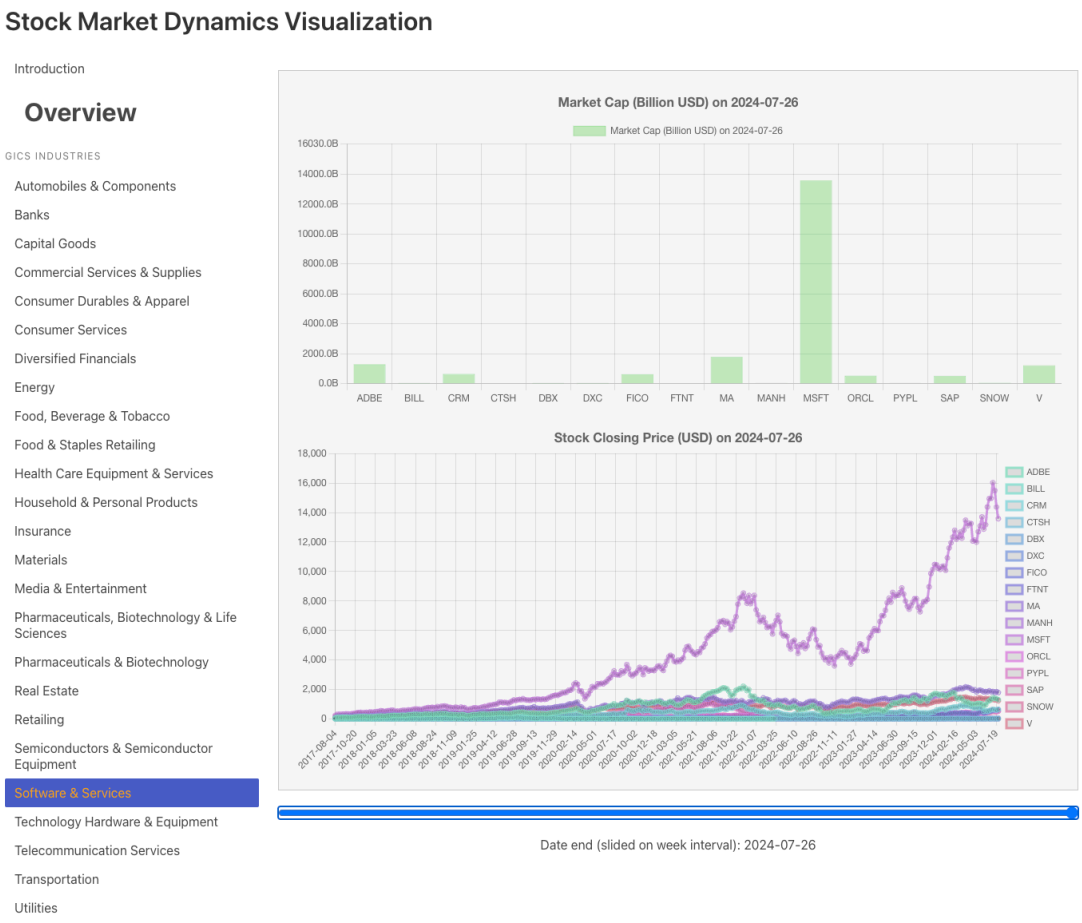


Figure 8.3: Industry-Specific Analysis View

8.2 Evaluation

8.2.1 Key Findings

Through our visualization system, we were able to answer our initial research questions:

1. Industry Evolution Patterns:

- The Software & Services industry showed remarkable growth, as evidenced by the increasing bubble sizes of companies like Microsoft
- Traditional industries like Utilities maintained relatively stable market capitalizations over time
- The Technology Hardware sector exhibited high concentration, with a few companies dominating the total market capitalization

2. Market Concentration:

- Within the Software & Services sector, we observed that market capitalization is heavily concentrated in a few major players (visible in Fig. 8.3)
- The bar chart comparison clearly shows the disparity between leading companies and others in the same sector

- Historical trends (shown in the line chart) reveal how this concentration has increased over time

3. Temporal Dynamics:

- The timeline feature revealed different growth patterns across industries
- Technology sectors showed more volatile but generally upward trends
- Traditional sectors demonstrated more stable, gradual growth patterns

8.2.2 Visualization Effectiveness

Our visualization system demonstrates several strengths:

- **Hierarchical Display:** Successfully shows both industry-level and individual company performance
- **Interactive Features:** Timeline slider and industry selection enable detailed exploration
- **Multi-view Analysis:** Combines overview patterns with detailed industry analysis

However, we also identified areas for improvement:

(a) **Scaling Challenges:**
   - Very large companies can sometimes dominate the visualization
   - Smaller companies might be difficult to identify in dense clusters

(b) **Information Density:**
   - The overview can become cluttered with too many data points
   - Additional filtering options could help manage complexity

(c) **Temporal Analysis:**
   - Could benefit from additional trend indicators
   - Might add annotations for significant market events

8.2.3 Future Improvements

Based on our evaluation, we have identified several promising directions for enhancing our visualization system. A key improvement would be the implementation of a logarithmic scale option, which would better handle the significant size disparities between major technology companies and smaller firms across various sectors. This would particularly benefit the overview visualization, where companies like Microsoft and Apple currently dominate the visual space.

To provide users with more focused analysis capabilities, we envision adding advanced filtering options that would allow them to concentrate on specific market capitalization ranges. This would enable more granular analysis of mid-cap and small-cap companies that are currently overshadowed by larger corporations. Additionally, we plan to incorporate industry-specific performance metrics to provide deeper insights into sector-level trends and patterns.

The temporal aspect of our visualization could be enhanced by adding historical event markers that would provide important context for market changes. For instance, marking significant events like major product launches, regulatory changes, or economic milestones would help users better understand the factors driving market capitalization changes.

Finally, we aim to develop additional visualization views specifically designed for cross-industry comparison, which would help users better understand the relationships and interactions between different sectors of the economy. These improvements would build upon our current framework while expanding its analytical capabilities and user experience.

This page is intentionally left blank.

Bibliography

[1] MSCI Inc. and S&P Global. Global Industry Classification Standard (GICS®), 1999.

[2] State Street Global Advisors. GICS Sector and Industry Map, 2024.

[3] MSCI Inc. Global Industry Classification Standard (GICS®), 2024.